Review Article

REVISTA BRASILEIRA DE FUTEBOL
The Brazilian Journal of Soccer Science

ISSN: 1983-7194

Predicting match outcomes in the 2006 football world cup considering time-effect weights

Previsão dos resultados dos jogos da Copa do mundo de futebol de 2006

Salasar, LEB[1]; Suzuki, AK[2] ; Louzada, F[2] ; Leite, JG[1]

1 – Department of Statistics, Universidade Federal de São Carlos, São Carlos, SP - Brasil

2 - Department of Applied Mathematics and Statistics, Universidade de São Paulo, São Carlos, SP - Brasil

Abstract

In this paper we propose a bayesian subjective statistical methodology for predicting match outcomes of 2006 World Cup Association Football. The prior information considered by the model is the opinion of experts and the FIFA scores announced previously to the start of the competition. As matches occur, the observed results are included in the model with different weights, which are inversely proportional to the elapsed time of the observation. In addition to the weights given to previous results, a weight is associated with the opinions of experts. This approach allows for the model calibration, directed to increase the predictive capability, through the appropriate choice of the weights associated with the experts' opinions and with the previous matches. The win, draw and loss probabilities for each match were obtained exactly and, by means of a stochastic simulation, we estimated the probabilities of classification in the group stage, to reach each round of the knockout stage and winning the tournament.

Keywords: Bayesian Inference, Simulation, Sport forecasting, World Cup Association Football.

Correspondência:

Luis Ernesto Bueno Salasar

Departamento de Estatística (DES) - Universidade Federal de São Carlos (UFSCar)

Rodovia Washington Luís, km 235 - SP-310

São Carlos, São Paulo - Brasil

CEP 13565-905

E-mail: luis@ufscar.br

Review Article

Resumo

Neste artigo propomos uma metodologia estatística bayesiana com enfoque subjetivista para previsão dos resultados dos jogos da Copa do mundo de futebol de 2006. A informação a priori considerada pelo modelo é a opinião de especialistas e os escores FIFA divulgados previamente ao início da competição. À medida que os jogos ocorrem, os resultados observados são incluídos na modelagem com diferentes pesos, que são inversamente proporcionais ao tempo decorrido da observação. Além dos pesos dados aos resultados anteriores, um peso é associado às opiniões dos especialistas. Esta abordagem permite a calibração do modelo, no sentido de aumentar a capacidade preditiva, por meio da escolha apropriada dos pesos associados à opinião dos especialistas e aos jogos anteriores. As probabilidades de vitória, empate e derrota em cada jogo foram obtidas de maneira exata e, a partir de uma simulação estocástica, estimamos as probabilidades de classificação na fase de grupo, de se chegar a cada uma das rodadas da fase eliminatória e de vencer o torneio.

Palavras chave: Copa do mundo de futebol, Inferência bayesiana, Previsão esportiva, Simulação.

Introduction

The World Cup tournament organized by FIFA is probable the most important international soccer championship. It takes place every four years, joining 32 teams from around the world, and is composed by two stages: a group stage followed by a knockout stage. In the group stage, teams compete within eight groups of four teams each. Each group plays a round-robin tournament and the top two teams from each group advance to the next stage. Points are assigned to each team within a group, where a win counts for three points and a draw counts for one. The teams are ranked on the following criteria in order: greatest number of points, greatest total goal difference, and greatest number of goals scored. If teams remain level after applying these criteria, a mini-group is formed with these teams and the same criteria applied. If after this teams remain level, a drawing of lots is held. The knockout stage is a single-elimination tournament in which teams play each other in one-off matches, with extra time of 30 minutes (2 halves of 15 minutes each) and penalty shootouts used to decide the winners if necessary.

In the soccer modeling literature, few articles can be found concerning score predictions for the World Cup. This can be explained by the limited amount of valuable data related to international matches due to great changes in national squads in the large elapsed time between World Cups (4 years), and also due to the fact that few competitions join teams from different continents.

Dyte and Clark[3] presents a log-linear Poisson regression model for soccer match predictions applied to the 1998 World Cup tournament, which takes the FIFA ratings as covariates. In that paper, the authors give some results about the predictive power of the model and also present simulation results to estimate winning championship probabilities.

Taking a different approach, Brillinger[1] proposed to model directly the win, draw and loss probabilities. In that paper, Brillinger employed a trinomial model and applied it to the Brazilian 2006 Series A championship to obtain a probability estimate of any particular team's being champion, fit the team's final points and to evaluate the chance of a team's being in the top four places.

Recently, Karlis and Ntzoufras[5] have applied the Skellam's distribution to model the goal difference between home and away teams. The authors argue that this approach does not rely neither on independence nor on the marginal Poisson distribution assumptions for the number of goals scored by the teams. A bayesian analysis for predicting match outcomes for the English Premiere League (2006-2007 season) is carried out using a log-linear link function and non-informative prior distributions for parameters.

Using a counting processes approach, Volf[7] modeled the development of a match score as two interacting time-dependent random point processes. The interaction between teams is modeled via a semi-parametric multiplicative regression model of intensity. The author has applied this model to the analysis of the performance of the eight teams that reached the quarter-finals of 2006 World Cup.

Suzuki et al.[6] proposes a Bayesian simulation methodology for predicting match outcomes applying it to the 2006 World Cup tournament (WCT). The authors have considered a Bayesian approach to predict the outcomes taking into account experts' opinion as prior information and the FIFA ratings as a covariate.

In this paper, we extend the Suzuki et al.[6] modeling by incorporating time-effect weights for the matches, that is, we consider that outcomes of matches which were played first have less importance than the outcomes of more recent matches. As an advantage of our approach it is possible to calibrate the experts' opinions as well as the importance of previous match outcomes in the modeling, directing for a control on the model prediction capability. Considering a grid of values for the experts' opinions weight a_0 and for the last match's importance, the p_i's values, we can assess the impact of these weights on the model prediction capability.

The paper is outlined as follows. In Section 2 we present the probabilistic model and expression for priors and posterior distribution of parameters, as well the predictive distributions. In Section 3 we present the method used to estimate the probabilities of winning the tournament and of reaching the final match, considering different values of weights for experts' opinion and past matches. In Section 4 we give our final considerations about the results and further work.

Probabilistic Model

In this section, following Suzuki et al.[6], we present the probabilistic model. Consider a match between teams A and B with respective FIFA ratings R_A and R_B. Assuming X_{AB} and X_{BA} the number of scored goals by team A and B, respectively, two independent random variables such that

$$X_{AB}|\lambda_A \sim Poisson\left(\lambda_A \frac{R_A}{R_B}\right), \qquad (1)$$

$$X_{BA}|\lambda_B \sim Poisson\left(\lambda_B \frac{R_B}{R_A}\right), \qquad (2)$$

where λ_A denotes the mean number of goals of the team A scored against team B, λ_B denotes the mean number of goals of the team B scored against team A, and R_A and R_B are the FIFA ratings for times A and B, respectively. In this model we use the FIFA ratings to quantify the ability for each team, such that the mean number of goals that team A scored against B is directly proportional to the rating of team A and inversely proportional to the rating of team B, and vice-versa.

Prior distribution

To formulate the prior distribution, a number of experts may give their guesses about the match's scores, instead of asking them directly for information about the parameters. Assuming the experts' opinions are independent and following a Poisson distribution, we shall obtain the prior distribution for the parameters using a procedure analogous to the power prior method Chen and Ibrahim[2] with the historical data replaced by the experts' opinion.

Previous to the knowledge of experts' opinions, we assume total absence of information, which will be expressed by the Jeffreys (noninformative) prior (Jeffreys[4]) for the Poisson model given by

$$\pi_{0(\lambda_A)} \propto \lambda_A^{-\frac{1}{2}} , \qquad (3)$$

$$\pi_{0(\lambda_B)} \propto \lambda_B^{-\frac{1}{2}} , \qquad (4)$$

Now, considering the experts' opinions (about the number of goals A scored against a team OA) $\tilde{x}^i_{A,OA}$, *i* =*1, …, s*, a random sample of a Poisson distribution with parameter $\lambda_A \frac{R_A}{R_{OA}}$, the power prior of λ_A is expressed as

$$\pi_0(\lambda_A|\mathcal{D}_0) \propto \pi_0(\lambda_A)\left\{\prod_{i=1}^{s} exp\left(-\frac{\lambda_A R_A}{R_{OA}}\right)\left[\frac{\lambda_A R_A}{R_{OA}}\right]^{\tilde{x}^i_{A,OA}}\right\}^{a_0}$$

$$\propto exp\left(-\frac{a_0 s \lambda_A R_A}{R_{OA}}\right)\lambda_A{}^{a_0\sum_{i=1}^{s}\tilde{x}^i_{A,OA}-\frac{1}{2}}, \qquad (5)$$

where $0 \leq a_0 \leq 1$ and $\mathcal{D}_0$ denotes all the experts' opinions.

So, denoting $\sum_{i=1}^{s}\tilde{x}^i_{A,OA}$ by $\tilde{x}_{A,OA}$, it follows from (5) that the power prior distribution of λ_A when $a_0 > 0$ is given by

$$\lambda_A|\mathcal{D}_0 \sim Gamma\left(a_0\tilde{x}_{A,OA}+\frac{1}{2}, a_0 s\frac{R_A}{R_{OA}}\right), \qquad (6)$$

and it is given by the Jeffreys prior (3) when $a_0 = 0$.

Analogously for team B, λ_B has power prior given by

$$\lambda_B|\mathcal{D}_0 \sim Gamma\left(a_0\tilde{x}_{B,OB}+\frac{1}{2}, a_0 s\frac{R_B}{R_{OB}}\right), \qquad (7)$$

when $a_0 > 0$ and the Jeffreys prior (4) when $a_0 = 0$.

Posterior and predictive distributions

Our interest is to predict the number of goals of the team *A* scored against team *B*, using all the available information (hereafter denoted by $\mathcal{D}$). This information is originated from two sources: the experts' opinions and the scores of matches already played. So, we may be in two distinct situations: (i) we do have the experts' opinions but no matches have been played; (ii) we have both the experts' opinions and the scores of played matches.

Following Suzuki et al.[6], in situation (i), we do not have observed data, only the experts' opinions. So, from the model given in (1) and the prior distribution in (6), it follows that the prior predictive distribution of X_{AB} is given by

$$X_{AB} \sim NB\left(a_0\tilde{x}_{A,OA}+\frac{1}{2}, \left[1+\frac{R_{OA}}{R_B a_0 s}\right]^{-1}\right) \qquad (8)$$

where NB denotes the Negative Binomial distribution.

Analogously for team *B*, from model (2) and the prior distribution (7), it follows that the prior predictive distribution of X_{BA} is

$$X_{BA} \sim NB\left(a_0\tilde{x}_{B,OB}+\frac{1}{2}, \left[1+\frac{R_{OB}}{R_A a_0 s}\right]^{-1}\right) \qquad (9)$$

In situation (ii), assume that team *A* has played *k* matches, the first against team C_1, the second against team C_2, and so on until the *k*-th match against team C_k. Moreover, suppose that, given λ_A, $X_{A,C_1}, \ldots, X_{A,C_k}$ are independent Poisson distributed random variables with parameters $\lambda_A\frac{R_A}{R_{C_1}}, \ldots, \lambda_A\frac{R_A}{R_{C_k}}$. Henceforth, we consider the weight p_i for the *i*-th outcome, $0 \leq p_i \leq 1$, *i* =1, …, *k*. So, from the model (1), the weighted likelihood is given by

$$L^p(\lambda_A|\mathcal{D}) = \prod_{i=1}^{k} P\left[X_{A,C_i} = x_A^i\right]^{p_i} \qquad (10)$$

$$\prod_{i=1}^{k}\frac{e^{-\lambda_A p_i\frac{R_A}{R_{C_i}}}\left(\lambda_A\frac{R_A}{R_{C_i}}\right)^{x_A^i p_i}}{\left[\left(x_A^i\right)!\right]^{p_i}}$$

$$\propto exp\left\{-\lambda_A\sum_{i=1}^{k}p_i\frac{R_A}{R_{C_i}}\right\}\lambda_A{}^{\sum_{i=1}^{k}x_A^i p_i} \qquad (11)$$

where $x_A^i = 0, 1, \ldots$ with $i = 1, \ldots, k$.

So, from the likelihood (11) and the prior distribution (6), it follows that the posterior distributions of parameter λ_A is

$$\lambda_A|\mathcal{D} \sim Gamma\left(a_0\tilde{x}_{A,OA} + \sum_{i=1}^{k} p_i x_A^i \right. \\ \left. + \frac{1}{2}, \sum_{i=1}^{k} p_i \frac{R_A}{R_{C_i}} + a_0 s \frac{R_A}{R_{OA}}\right). \qquad (12)$$

Thus, from the model (1) and the posterior (12), the posterior predictive distribution of X_{AB} is

$$X_{AB}|\mathcal{D} \sim NB\left(a_0\tilde{x}_{A,OA} + \sum_{i=1}^{k} p_i x_A^i \right. \\ \left. + \frac{1}{2}, \frac{\sum_{i=1}^{k} \frac{p_i}{R_{C_i}} + \frac{a_0 s}{R_{OA}}}{\sum_{i=1}^{k} \frac{p_i}{R_{C_i}} + \frac{a_0 s}{R_{OA}} + \frac{1}{R_B}}\right). \qquad (13)$$

Analogously for team *B*, the posterior distribution of λ_B is

$$\lambda_B|\mathcal{D} \sim Gamma\left(a_0\tilde{x}_{B,OB} + \sum_{i=1}^{k} p_i x_B^i \right. \\ \left. + \frac{1}{2}, \sum_{i=1}^{k} p_i \frac{R_B}{R_{D_i}} + a_0 s \frac{R_B}{R_{OB}}\right). \qquad (14)$$

where D_i, $i = 1, \ldots, k$, are the opponent teams faced by *B*.

Hence, from the model (2) and the posterior (14), the posterior predictive distribution of X_{BA} is

$$X_{BA}|\mathcal{D} \sim NB\left(a_0\tilde{x}_{B,OB} + \sum_{i=1}^{k} p_i x_B^i \right. \\ \left. + \frac{1}{2}, \frac{\sum_{i=1}^{k} \frac{p_i}{R_{D_i}} + \frac{a_0 s}{R_{OB}}}{\sum_{i=1}^{k} \frac{p_i}{R_{D_i}} + \frac{a_0 s}{R_{OB}} + \frac{1}{R_A}}\right). \qquad (15)$$

By considering the posterior predictive distributions (13) and (15) we are able to answer the following question. Suppose a team reaches the quarterfinals round, should we give the same importance to the matches of the group stage and to that match of the round of sixteen? If the answer is affirmative, it suffices to choose all the p_i values equal to 1 to make all the matches equally important and the posterior predictive distributions (13) and (15) are identical to those presented at Suzuki et al[6]. However, if the answer is negative, the p_i values should be chosen according to the relative importance of the previous matches.

Tournament Simulation

In this section, we perform a simulation of the whole competition in order to estimate probabilities of winning the tournament and reaching the final match according to different weights for the past matches, p_i's, and different weight for the experts' opinions, a_0. Just prior to each of the last 5 rounds, one simulation of 10,000 tournament replicas has been carried out for each studied situation and the probability of a given event is estimated by its proportion of occurrence over all tournament replicas.

General Specifications

Through the predictive distributions we can randomly generate a score for each match and so, repeating this procedure for all matches, we can simulate a replica of the whole competition. Considering a grid of values for the experts' opinions weight a_0 and for the last matches' importance, the p_i's values, we can assess the impact of these weights on the model prediction capability.

For a given match played by teams *A* and *B*, we calculate the probabilities of win (P_W), draw (P_D) and loss (P_L) of team *A* from the predictive distributions (13) and (15), using the following expressions,

$$P_W = P(X_{AB} > X_{BA}) \\ = \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} P(X_{AB}=i)\, P(X_{BA}=j), \qquad (16)$$

$$P_D = P(X_{AB} = X_{BA}) = \sum_{i=0}^{\infty} P(X_{AB} = i)\,P(X_{BA} = i), \quad (17)$$

$$P_L = P(X_{AB} < X_{BA}) = \sum_{j=1}^{\infty}\sum_{i=0}^{j-1} P(X_{AB} = i)\,P(X_{BA} = j). \quad (18)$$

Just before each round, 10 experts give their opinions about the scores of all matches in that round. The a_0 represents the level of confidence on the experts' opinion. For $a_0 = 0$ the experts' opinion is disregarded. To account for the mean experts' opinion, we have chosen $a_0 = 1/10$, in the sense that, if one observation equal to the mean of the experts' opinion is taken from the sampling distribution, then under the noninformative Jeffreys priors the posterior distribution is the same as the power prior distributions (6) and (7). It is important to note that it is not allowed for the experts to discuss about their guesses in order to make the guesses as much independent as possible. Also, the opinions only refer to one round at a time and they may be influenced by the outcomes of matches already played. Observed that if $a_0 = 1$ we give full confidence to the experts' opinion, on the other hand if $a_0 < 1$ the level of confidence is decreased. We have considered the values 0, 0.1, 0.5, 0.9, 1 for a_0.

The p_i's values represent the importance attached to the past matches. For instance, if a team has reached the semifinal round and so there is 5 previous matches observed for this team, we will have to set 5 weights (p_1, p_2, p_3, p_4, p_5) for those matches. Different choices of the p_i's values were taken including the one where all the weights equal to 1.

Results

In this section we display the probabilities of winning the tournament and of reaching the final match considering the observed data available before the start of the round of sixteen, the quarterfinals and the semifinal round. In addition, we present the estimated probabilities for the final match Italy vs France.

Tables 1 and 2 present the estimated probabilities of winning the tournament and of reaching the final match, respectively, for the top eight teams of the tournament only considering the available information before the round of sixteen. Table 1 shows that Brazil is the team with highest probabilities of winning the tournament despite of the choice of the weights, except for the choice (p_1, p_2, p_3) = (1, 1, 1). For this choice, Argentina has the highest probabilities. We also observe that increasing the weight of the experts' opinion leads to an increase in the winning probability for five of the semifinalist teams (Argentina, Italy, France, Germany and Portugal), indicating that the use of experts information improve the prediction capability. On the other hand, for Portugal and Ukraine teams, the winning probability decreases as the experts' weight increase, that is, experts' opinion were unfavorable for those teams. Moreover, setting high values for the weight of the most recent matches leads to an increase in the probabilities of reaching the final match for Germany and Argentina teams.

From the results presented at Table 2, we conclude that Brazil and Germany teams have the greatest probabilities of reaching the final match. As the experts' weight increase, the probabilities of reaching the final match increases for Italy, France, Germany, Brazil and Argentina. For Portugal and Ukraine, the latter probability decreases as the experts' weight increase.

Table 1. Estimated probabilities (%) of winning the tournament just before the round of sixteen.

		a_0				
	(p_1, p_2, p_3)	**0**	**0.1**	**0.5**	**0.9**	**1.0**
Italy	(0.10,0.25,1.00)	7.39	7.89	7.69	7.53	7.51
	(0.25,0.50,1.00)	6.53	7.67	7.45	7.44	7.2
	(0.50,0.75,1.00)	6.23	6.35	6.89	6.89	6.7
	(0.80,0.90,1.00)	5.81	6.38	6.69	6.78	6.41
	(1.00,1.00,1.00)	5.46	5.75	6.56	6.94	6.87
France	(0.10,0.25,1.00)	4.26	4.54	5.81	6.17	6.34
	(0.25,0.50,1.00)	3.08	3.81	4.82	5.37	5.46
	(0.50,0.75,1.00)	1.67	2.32	3.90	5.00	4.73
	(0.80,0.90,1.00)	1.07	1.64	3.58	4.08	4.56
	(1.00,1.00,1.00)	0.73	1.15	3.22	4.29	3.99
Germany	(0.10,0.25,1.00)	13.28	15.00	16.15	15.75	16.4
	(0.25,0.50,1.00)	13.34	13.7	16.12	16.08	16.77
	(0.50,0.75,1.00)	13.31	14.83	16.05	15.96	16.31
	(0.80,0.90,1.00)	15.47	15.3	16.96	17.27	16.87
	(1.00,1.00,1.00)	15.78	16.66	17.19	16.63	16.78
Portugal	(0.10,0.25,1.00)	9.54	8.00	4.79	4.45	4.21
	(0.25,0.50,1.00)	8.34	6.90	5.19	4.59	4.07
	(0.50,0.75,1.00)	7.43	6.64	4.98	4.65	3.78
	(0.80,0.90,1.00)	7.15	5.96	4.64	4.27	3.92
	(1.00,1.00,1.00)	5.91	5.50	4.49	4.27	4.31
Brazil	(0.10,0.25,1.00)	24.63	25.6	24.24	23.48	23.23
	(0.25,0.50,1.00)	23.66	24.17	23.60	23.13	22.87
	(0.50,0.75,1.00)	21.06	21.37	22.44	21.99	22.3
	(0.80,0.90,1.00)	18.85	19.19	21.19	21.05	21.58
	(1.00,1.00,1.00)	16.54	18.46	20.83	22.09	21.66
Argentina	(0.10,0.25,1.00)	4.17	8.01	16.91	20.37	20.43
	(0.25,0.50,1.00)	8.89	13.55	19.13	21.92	22.83
	(0.50,0.75,1.00)	14.48	17.78	22.47	24.42	24.12
	(0.80,0.90,1.00)	16.67	20.06	23.18	24.85	25.18
	(1.00,1.00,1.00)	18.54	21.53	24.61	24.38	25.56
England	(0.10,0.25,1.00)	6.86	7.36	7.81	6.84	6.65
	(0.25,0.50,1.00)	5.62	5.93	6.57	6.09	5.75
	(0.50,0.75,1.00)	4.02	4.44	5.08	5.60	6.11
	(0.80,0.90,1.00)	2.86	3.70	4.78	4.90	5.36
	(1.00,1.00,1.00)	2.54	3.02	4.35	4.74	4.91
Ukraine	(0.10,0.25,1.00)	3.07	2.76	2.20	1.98	1.81
	(0.25,0.50,1.00)	3.86	3.22	2.22	2.08	2.07
	(0.50,0.75,1.00)	4.22	3.55	2.49	1.94	2.13
	(0.80,0.90,1.00)	3.18	3.28	2.15	1.93	2.17
	(1.00,1.00,1.00)	3.07	2.92	2.12	1.92	1.91

Table 2. Estimated probabilities (%) of reaching the final match just before the round of sixteen.

		a_0				
	(p_1,p_2 p_3)	**0**	**0.1**	**0.5**	**0.9**	**1.0**
Italy	(0.10,0.25,1.00)	15.79	17.34	17.84	18.01	17.49
	(0.25,0.50,1.00)	14.57	16.78	17.40	17.35	16.94
	(0.50,0.75,1.00)	13.83	14.48	16.06	16.08	16.15
	(0.80,0.90,1.00)	13.01	14.25	15.24	15.60	15.24
	(1.00,1.00,1.00)	12.69	13.21	15.02	16.05	15.81
France	(0.10,0.25,1.00)	8.80	9.68	12.75	13.40	12.94
	(0.25,0.50,1.00)	6.66	8.06	10.66	11.98	12.19
	(0.50,0.75,1.00)	4.45	5.69	9.64	11.49	11.42
	(0.80,0.90,1.00)	3.09	4.51	8.99	9.79	10.65
	(1.00,1.00,1.00)	2.09	3.58	7.90	9.53	9.95
Germany	(0.10,0.25,1.00)	24.99	27.37	28.17	27.95	28.22
	(0.25,0.50,1.00)	24.53	25.17	27.53	27.63	27.72
	(0.50,0.75,1.00)	24.78	26.00	27.25	27.25	27.25
	(0.80,0.90,1.00)	27.49	27.10	28.68	28.57	28.36
	(1.00,1.00,1.00)	27.60	28.10	29.20	28.31	28.37
Portugal	(0.10,0.25,1.00)	18.36	15.66	12.00	11.19	11.66
	(0.25,0.50,1.00)	16.57	14.79	12.56	11.53	11.45
	(0.50,0.75,1.00)	16.34	15.00	12.63	11.88	10.95
	(0.80,0.90,1.00)	15.70	13.94	12.47	11.69	11.0
	(1.00,1.00,1.00)	13.84	13.38	11.84	11.48	11.8
Brazil	(0.10,0.25,1.00)	35.47	37.10	37.48	37.85	37.75
	(0.25,0.50,1.00)	34.57	36.42	37.20	38.23	38.41
	(0.50,0.75,1.00)	32.55	34.04	36.60	36.98	36.94
	(0.80,0.90,1.00)	30.78	32.02	35.79	36.47	36.96
	(1.00,1.00,1.00)	28.61	31.63	35.89	37.25	36.76
Argentina	(0.10,0.25,1.00)	10.52	16.97	29.74	34.33	34.38
	(0.25,0.50,1.00)	18.33	24.89	32.61	35.68	36.4
	(0.50,0.75,1.00)	26.21	30.77	35.66	38.45	38.43
	(0.80,0.90,1.00)	28.78	32.81	36.82	38.61	38.85
	(1.00,1.00,1.00)	31.14	34.66	37.93	38.08	39.02
England	(0.10,0.25,1.00)	14.10	15.18	16.71	15.64	15.84
	(0.25,0.50,1.00)	12.26	13.25	14.92	15.06	15.02
	(0.50,0.75,1.00)	9.26	10.56	12.82	14.16	14.67
	(0.80,0.90,1.00)	7.48	9.90	12.41	13.28	14.03
	(1.00,1.00,1.00)	7.06	8.82	11.68	12.98	13.58
Ukraine	(0.10,0.25,1.00)	7.78	8.11	6.42	5.31	5.22
	(0.25,0.50,1.00)	9.63	8.46	6.67	6.00	5.46
	(0.50,0.75,1.00)	10.2	8.86	6.73	5.73	5.52
	(0.80,0.90,1.00)	8.51	8.15	6.03	5.42	5.63
	(1.00,1.00,1.00)	8.67	7.63	5.95	5.77	5.19

Table 3 shows very similar probabilities of reaching the final match over all choices of the weights, except for Ukraine whose probability have a great decrease if the weight a_0 is greater than 0.5, expressing the low confidence of experts in Ukraine qualification for semifinals. In fact, Italy beat Ukraine by 3 – 0 in quarterfinals. On the other hand, the chance of reaching the final match increases for Brazil, Germany and Portugal teams if the weight a_0 is greater. We also observe that different choices for the pi's weights for past matches do not alter significantly the results.

Table 3. Estimated probabilities of reaching the final match just before quarter-finals.

				a_0		
	(p_1,p_2,p_3, p_4)	**0**	**0.1**	**0.5**	**0.9**	**1.0**
Italy	(0.10,0.25,0.50,1.00)	13.23	12.33	11.16	10.73	10.59
	(0.25,0.50,0.75,1.00)	13.56	12.16	10.94	10.00	10.70
	(0.70,0.80,0.90,1.00)	13.21	12.71	11.31	10.87	10.05
	(1.00,1.00,1.00,1.00)	14.04	12.67	11.55	10.79	10.04
France	(0.10,0.25,0.50,1.00)	12.06	11.75	12.40	12.88	12.90
	(0.25,0.50,0.75,1.00)	11.74	11.70	12.81	12.48	12.56
	(0.70,0.80,0.90,1.00)	12.07	12.24	12.61	12.74	13.15
	(1.00,1.00,1.00,1.00)	11.72	12.47	13.00	12.75	13.07
Germany	(0.10,0.25,0.50,1.00)	11.94	13.28	16.75	17.97	17.86
	(0.25,0.50,0.75,1.00)	11.42	12.70	16.46	17.77	18.69
	(0.70,0.80,0.90,1.00)	11.05	12.80	15.51	16.82	17.38
	(1.00,1.00,1.00,1.00)	10.65	11.92	15.03	16.89	17.31
Portugal	(0.10,0.25,0.50,1.00)	13.66	15.09	17.15	18.01	18.53
	(0.25,0.50,0.75,1.00)	12.94	14.81	16.87	18.47	17.52
	(0.70,0.80,0.90,1.00)	13.66	14.86	17.46	17.74	17.28
	(1.00,1.00,1.00,1.00)	14.26	15.15	16.42	17.30	17.53
Brazil	(0.10,0.25,0.50,1.00)	13.96	15.63	17.92	18.67	18.42
	(0.25,0.50,0.75,1.00)	15.32	16.23	17.98	19.12	18.82
	(0.70,0.80,0.90,1.00)	14.94	15.87	17.73	19.08	19.72
	(1.00,1.00,1.00,1.00)	14.61	15.89	18.30	18.62	19.77
Argentina	(0.10,0.25,0.50,1.00)	13.72	14.75	14.13	13.33	13.48
	(0.25,0.50,0.75,1.00)	14.58	14.89	14.32	13.18	13.05
	(0.70,0.80,0.90,1.00)	15.00	15.19	13.61	13.37	13.82
	(1.00,1.00,1.00,1.00)	14.88	14.33	14.58	13.92	13.12
England	(0.10,0.25,0.50,1.00)	11.78	10.49	8.33	7.23	7.32
	(0.25,0.50,0.75,1.00)	11.29	11.08	8.25	7.66	7.46
	(0.70,0.80,0.90,1.00)	11.11	10.12	9.12	7.76	7.16
	(1.00,1.00,1.00,1.00)	11.39	11.06	8.47	8.07	7.75
Ukraine	(0.10,0.25,0.50,1.00)	9.65	6.68	2.16	1.18	0.90
	(0.25,0.50,0.75,1.00)	9.15	6.43	2.37	1.32	1.20
	(0.70,0.80,0.90,1.00)	8.96	6.21	2.65	1.62	1.44
	(1.00,1.00,1.00,1.00)	8.45	6.51	2.65	1.66	1.41

Figure 1 displays the probabilities of reaching the final match for the top 4 teams (Italy, France, Germany and Portugal) calculated before the semifinals took place. It is remarkable the increasing of the Portugal and the decreasing of France probabilities with the increasing of the a_0 weight, indicating a high level of confidence of experts on a Portugal win against France in the semifinals, which did not occur. Further, we observe that increasing the importance of the most recent matches leads to greater probabilities of Germany win against Italy in the semifinals, which is explained by the most expressive victories of Germany on the recent matches (beat Sweden in the round of sixteen and Argentina in the quarterfinals) in comparison with Italy (beat Switzerland in the round of sixteen and Ukraine in the quarterfinals).

Figure 1. Estimated probabilities of reaching the final match (just before the semifinals).

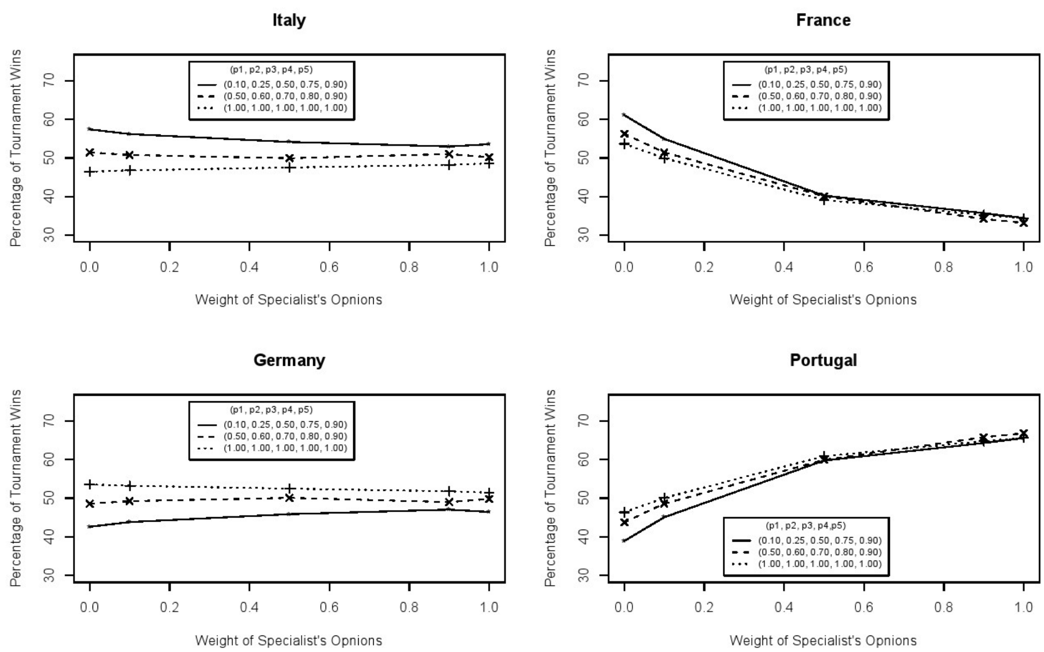


Figure 2 presents the winning tournament probabilities for the top 4 teams (Italy, France, Germany and Portugal) before the start of the semifinals. The same previous comments relative to Figure 1 also apply in this case.

Figure 3 displays the winning probabilities for the final match Italy versus France. It is worth noting that with no expert information ($a_0 = 0$), the winning probabilities for both teams are close to 50%, indicating that previous performance of both teams are quite similar. As experts' opinion weight a_0 increases, the probability of France victory increases, reflecting the experts' opinion in favor of France. Assigning greater relative weight for the most recent matches, the winning probabilities of France increase, which can be explained by the stronger opponents defeated by France in knockout stage matches: France defeated Spain, Brazil and Portugal, while Italy defeated Australia, Ukraine and Germany.

Figure 2. Estimated winning tournament probabilities (just before the semifinals).

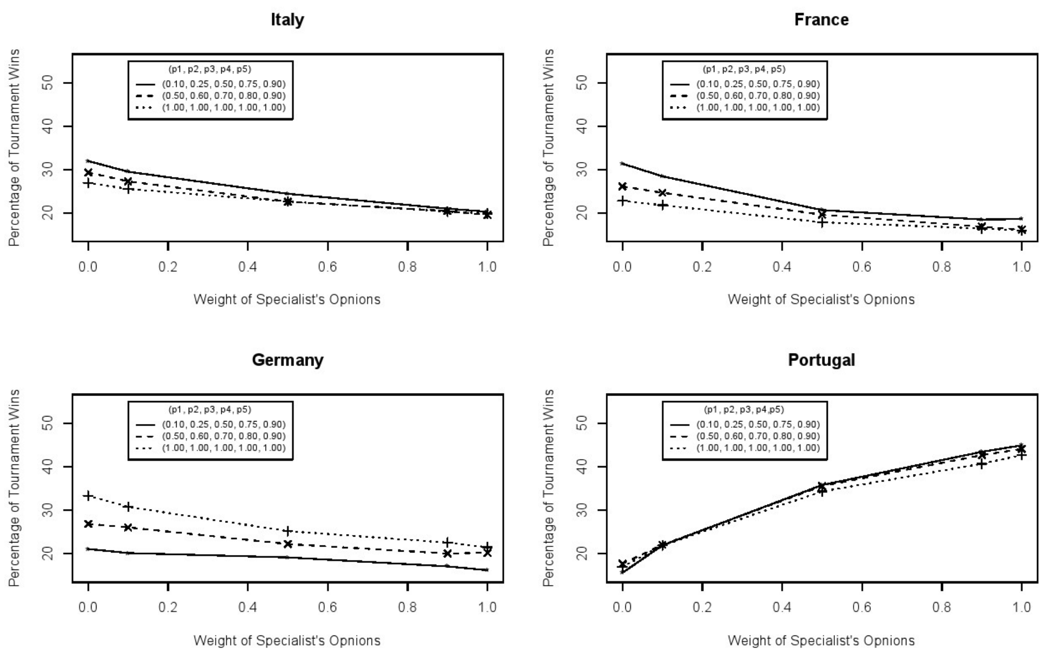
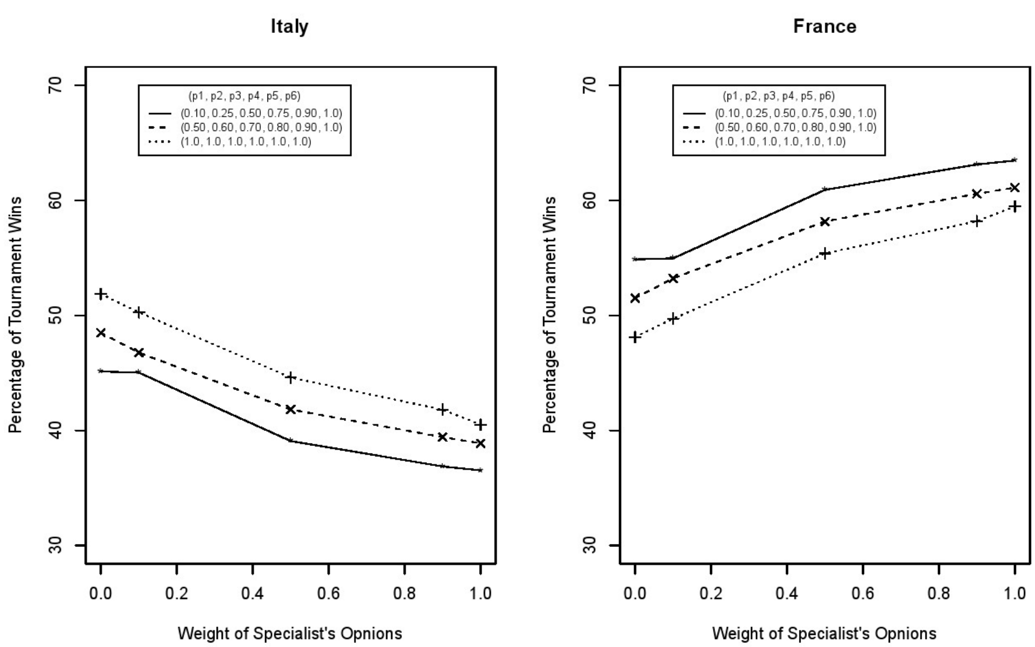


Figure 3. Estimated winning probabilities before the final match.



Final Remarks

In this paper we present an extension of the model proposed by Suzuki et al.[6] by incorporating time-effect weights for the past matches, which is very intuitive in practice and allows for model calibration in order to obtain more accurate predictions. The prior distributions were updated every round, i. e., experts were consulted after the finish of all previous matches and before the beginning of the match of interest, providing great flexibility to the modeling since experts' opinion were updated with all the previous information. Moreover, the calculation of predictions using our model is very simple because all the predictive distributions have a closed analytical form. This makes the random generation process faster and, consequently, simplifies the task of calculating the probabilities of interest.

From the practical point of view, the present study gives a better idea about how the previous matches and experts' opinion influence the estimated probabilities of the events. In the present model, the weights for the past matches and for the experts' opinions are fixed over the study and all experts have the same weight value. A future development of this work would be allowing for different weights for each expert in order to take into account different levels of knowledge among experts panel. Further, it would be useful to let the experts' weights be updated as the competition progresses based on the accuracy of the predictions given by each expert. Another model improvement would be to assume a hierarchical structure for the weights of the past matches, assuming a given parametric distribution for the weights depending on fixed hyperparameters. The same hierarchical idea would apply to the expert weight, providing a full hierarchical structure specification, as suggested by Chen and Ibrahim[2] in the context of the power prior distributions.

References

1. Brillinger DR. Modelling game outcomes of the Brazilian 2006 series a championship as ordinal-valued. Braz J Probab Stat. 2008; 22(2):89-104.
2. Chen MH, Ibrahim, JG. Power prior distributions for regression models. Stat Sci. 2000; 15(1):46-60.
3. Dyte D, Clarke SR. A ratings based Poisson model for world cup soccer simulation. J Oper Res Soc. 2000; 51(8):993-998.
4. Jeffreys H. Theory of Probability. 3rd ed. Oxford: Clarendon Press; 1961.
5. Karlis D, Ntzoufras I. Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference. IMA J Manag Math. 1961; 20:133-145.
6. Suzuki AK, Salasar LEB, Louzada-Neto F, Leite JG. A bayesian approach for predicting match outcomes: The 2006 (association) football world cup. J Oper Res Soc. 2009; 61:1530-1539.
7. Volf P. A random point process model for the score in sport matches. IMA J Manag Math. 2009; 20:121-131.

b